Database
(2016/07/01)

[1]

Consider a fictitious DBMS, called F-DBMS that uses a relational approach and supports a simple SQL-like language. The overview of this fictitious DBMS is as follows:

F-DBMS uses a basic file system (such as the one in UNIX or Linux) to store relations. There is one file for each relation. For example, there is a file named R for relation R. The file for a relation has one line for each tuple. Values of attributes of a tuple are stored as character strings, separated by two consecutive special marker characters ## that are assumed not to appear in the data. The DB schema is stored in a special file named `schemas`. For each relation, the file `schemas` has a line beginning with that relation name, in which attribute names alternate with types. The two consecutive characters ## separate these elements.

To process a SQL-like query such as `SELECT A1, A2 FROM R WHERE <Condition>`,
(i) F-DBMS reads file `schemas` to find out the schema of relation `R` in the `FROM` clause (ii) Check that the query is semantically valid for `R` (e.g., check that the name of the relation in the FROM clause exists in file `Schemas`, names of attributes in the `SELECT` clause are correct, etc), (iii) Read each line of the file named `R`, check `<Condition>` is satisfied, and output the values of attributes `A1` and `A2` in the line as a result tuple, if the condition is true.

Assume that grouping/ordering related clauses and insert/delete/update statements in SQL are also supported in some simple ways such as above. Assume also that embedding SQL statements in some programming language is also possible. Except for supporting SQL statements as described above, no other mechanism is provided in this DB system.

Describe possible problems (or inconvenience) of this fictitious DBMS for various database applications. (20 points)

[2]

(a) A set of relational operators is complete if any other relational operations can be expressed in terms of operators from this set. What are these operators? (5 points)

(b) Briefly describe the converting procedure from the ER-diagram to the relational schema discussed in the class. You are required to describe (i) the basic idea together with how to reduce the number of tables, and (ii) how to handle the weak entity set. (15 points)

[3]

(a) Briefly describe “dirty data” and also describe the risk when reading dirty data. (5 points)

(b) Briefly explain four isolation levels in the SQL standard. You have to describe main characteristics of each isolation level. (8 points)

(c) Explain what the update lock is.
Give the compatibility matrix for S(shared), X(exclusive) and U(update) locks. (7 points)

[4]

(a) Write a procedure that inserts a new data region *R* in the R-tree. (10 points)

(b) For relation `Suppliers(s-name, city, amount-supplied, item)`, we have six records numbered from 1 to 6 as shown below.

1: (S2, Chicago, 60, necklace) 2: (S1, Boston, 60, ring) 3: (S3, Dallas, 75, earring)
4: (S4, Dallas, 100, earring) 5: (S3, Dallas, 120, ring) 6: (S5, LA, 100, ring)

i) Give the bitmap index for “`city`” and the bitmap index for “`amount-supplied`”. (5 points)

ii) Now, let us delete the tuple (S4, Dallas, 100, earring), and then insert the tuple (S5, New York, 75, watch). Show your final bit map indexes after these insertion and deletion on the bit map indexes you just constructed for the above question “i)”. (5 points)

[5] Consider the Employee database below.

Employee (<u>e-name</u>, e-street, e-city)
Company (<u>c-name</u>, c-city)
Works (<u>e-name</u>, <u>c-name</u>, salary)
Manages (<u>e-name</u>, m-name)

Write the following queries in SQL:
Find the companies, which are located in New York, whose employees earn a higher salary on average than the average salary at ‘City Bank’. (20 points)